

Programma « AGORA »

Eindverslag – deel «Synthese van het onderzoek»

DOCDWH

UITBOUW EN METHODOLOGISCHE UITDIEPING VAN HET DATAWAREHOUSE ARBEIDSMARKT EN SOCIALE BESCHERMING

COÖRDINATOR:

PROF. J. BERGHMAN (CESO – KU LEUVEN)

PROMOTOR(EN):

PROF. J. BERGHMAN (CESO – KU LEUVEN), PROF. P. DESMAREZ (METICES – ULB), PROF. L. SELS (STEUNPUNT WSE – KU LEUVEN)

ONDERZOEKER(S):

HANS KNAPEN (CESO – KU LEUVEN), STIJN BRAES (STEUNPUNT WSE – KU LEUVEN), WIM HERREMANS (STEUNPUNT WSE – KU LEUVEN), THOMAS ERMANS (METICES – ULB)

DATUM: 20/03/2014

Deze synthese heeft tot doel de resultaten van het onderzoek op Internet weer te geven. Ze wordt door de coördinator van het netwerk aan de programma administrator, ter goedkeuring ***en in drie talen (Nederlands) , Frans en Engels), voorgelegd.***

Links naar de websites van het project

- http://www.ksz.fgov.be/nl/bcss/nodepage/content/websites/belgium/statistics/statistics_01.html

****

- NEDERLANDS -

Inleiding

Het project DOCDWH omvatte de inhoudelijke uitdieping en methodologische uitbouw van het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB) waarbij (1) het oorspronkelijke basispakket van data, afkomstig van de instellingen van de sociale zekerheid, en van metadata werd uitgediept en (2) aangevuld met nieuwe data en metadata.

Het project DOCDWH stond onder leiding van Prof. J. Berghman (Centrum voor Sociologisch Onderzoek – KU Leuven), die optrad als coördinator, en de promotoren Prof. P. Desmarez (METICES/TEF – Université Libre de Bruxelles) en Prof. L. Sels (Steunpunt WSE – KU Leuven). Het project, dat op aanvraag van de Kruispuntbank van de Sociale Zekerheid (KSZ) en de FOD Sociale Zekerheid (FOD SZ) werd uitgevoerd, liep van december 2009 tot februari 2014.

Context

Bij de realisatie van de twee centrale doelstellingen werd voortgebouwd op het AGORA-project DW-AMSB (AG/01/115) dat als doel had een grondige inhoudelijke analyse te maken van de bronnen opgenomen in het DWH AM&SB en een methodologische handleiding uit te werken. Dit project beoogde namelijk de ontsluiting van gegevensbronnen met het oog op de verbetering van de kwantitatieve rapportering, de analyse van beleidsvoorstellen en de uitvoering van onderzoek met betrekking tot de sociale bescherming door de FOD SZ of anderen. Bij de oplevering van dit project in 2009 werd de documentatie ter beschikking gesteld aan het grote publiek via de website van de KSZ, aan de hand van downloadbare documenten. Deze documentatie kan beschouwd worden als een basispakket.
Echter, het DWH AM&SB is geen statisch geheel. Wetswijzigingen, herschikkingen in de reglementeringen, reorganisaties binnen de sociale zekerheid, technische hervormingen aan databanken, veranderende variabelen, enzovoort hebben een invloed op de primaire en secundaire databanken waarop het DWH AM&SB is gebouwd. Deze gebeurtenissen kunnen ook rechtsreeks het DWH AM&SB beïnvloeden. Bovendien breidt het aantal bronnen gestaag uit. Al deze factoren hebben bijgevolg een grote impact op de documentatie en dienen permanent te worden opgevangen.

Doelstellingen en uitgevoerde taken

De twee centrale doelstellingen werden uitgesplitst in meerdere subdoelstellingen, namelijk:

(1) De mogelijkheden die het DWH AM&SB biedt voor het sociaal-economisch onderzoek en nationale en internationale rapportering moeten worden geanalyseerd. Hierbij moet aandacht worden besteed aan de vergelijking met andere bronnen voor socio-economisch onderzoek, de moeilijkheden en mogelijke valkuilen bij interpretatie van de data en de zwakten en sterkten van het DWH AM&SB.

(2) Er moeten een aantal nieuwe noties worden uitgewerkt, namelijk nieuwe afgeleide variabelen. Deze variabelen tonen de directe meerwaarde van het DWH AM&SB aan,

namelijk de koppeling van gegevens van verschillende bronnen.
(3) De impact van de wetgeving en de wijzing ervan op het DWH AM&SB moet verder worden uitgewerkt. Hiervoor moet een inventaris worden opgemaakt van de belangrijkste maatregelen die een impact hebben op de populaties, bronnen, variabelen en codes van het DWH AM&SB. Daarenboven moet worden nagegaan in welke mate de impact van dergelijke maatregelen kan worden gemeten aan de hand van het DWH AM&SB.

(4) De wijze waarop de data worden opgenomen in de primaire databanken moet verder worden gedocumenteerd. Dit is essentieel om een goed zicht te krijgen op de kwaliteit van de data.

(5) Er moet in de actualisering worden voorzien van de documentatie die is opgeleverd in het kader van de projecten DW-AMSB en 'integratie van de gegevens van de VDAB/FOREM/Actiris/ADG in het DWH AM&SB' (AGORA-project met referentie AG/00/143). Bovendien moet de gehele documentatie worden geïntegreerd in een metadatadocumenteringssysteem.

(6) Er dient te worden nagegaan in welke mate overheidsinstanties van verschillende departementen kunnen samenwerken voor het meedelen van data in het kader van socio-economische onderzoeksprojecten die het domein van één enkel overheidsdepartement overschrijden.

Om de vooropgestelde subdoelstellingen en resultaten te realiseren werden de werkzaamheden opgesplitst in elf taken en deeltaken. Deze (deel)taken worden weergegeven in onderstaande sectie bij de resultaten.

Resultaten

T1a: actualiseren van de documentatie en de handleiding die werd opgeleverd in het kader van de projecten AG/01/115 en AG/00/143
De fiches van RSZ, RSZPPO, RSVZ, VDAB-FOREM-Actiris-ADG, PDOS, CIMIRe/Sigedis, RVP, RVA, POD MI, FBZ, FOD SZ, RKW, RSVZ-kinderbijslagen, POD MI ESF, FAO, Rijksregister, minirn, RSVZ-loopbanen, herkomstgegevens, NIC, nomenclatuur, gebruikershandleiding, inkomen en de beschrijving van de voorwaarden voor de nomenclatuur werden permanent geactualiseerd. Het loopbaanbestand - werknemers dat in het verleden werd aangeleverd door CIMIRe wordt vanaf 2010 aangeleverd door Sigedis. Er werd nieuwe documentatie i.v.m. deze bron opgesteld.

T1b: aanvullen van de metadata met informatie over de wijze waarop de data worden opgenomen in de primaire databanken
Om de kwaliteit van de gegevens in het DWH AM&SB beter te kunnen inschatten is het noodzakelijk een zicht te hebben op de wijze waarop deze worden ingezameld. Ten eerste werd een template (fiche) opgesteld om deze informatie uniform te documenteren. Vervolgens werd in overleg met de betrokken instellingen voor iedere bron deze fiche ingevuld. Deze documentatie is beschikbaar voor de gebruikers via de website van de KSZ.

T1c: integreren van de bestaande en nieuwe documentatie in het verder te ontwikkelen Content Management System
Ten eerste werd, in overleg met Smals, de bestaande en nieuwe documentatie geïntegreerd in beConnected (een Alfresco-toepassing) in 2011. Dit online platform stroomlijnt de procedure (= workflow) waarbij er aanpassingen worden aangebracht in de documentatie (inhoudelijke wijzingen – vertalingen – publicatie). Het actualiseren en aanpassen van de gehele

documentatie verloopt sindsdien via dit platform. Hierbij hebben de onderzoekers via een e-id kaartlezer toegang tot beConnected.
Daarnaast werd de bestaande documentatie zowel vormelijk als inhoudelijk aangepast (benaming bestanden, nummering variabelen en codes, uniformiseren tabellen) als voorbereiding op de integratie van de metadata in het Content Management System (CMS).
Met de functionele uitwerking van het CMS werd reeds gestart tijdens het voorafgaandelijke project DW-AMSB. Echter, de verdere uitwerking van het CMS liep tijdens DOCDWH vertraging op, maar momenteel is alle documentatie opgeladen in het systeem. De onderzoeksploeg testte het CMS uitvoerig op consistentie, volledigheid, gebruiksvriendelijkheid,... Nadat de laatste 'bugs' door Smals zijn verholpen, zal het CMS aan de Gebruikersgroep worden gepresenteerd (dit is voorzien voor het tweede kwartaal 2014). Dit systeem vervangt de Word-documenten die worden aangeboden via de website van de KSZ. Het CMS maakt het mogelijk voor de gebruikers om alle documentatie inzake het DWH AM&SB op een gebruiksvriendelijkere manier te raadplegen. Zo kunnen variabelen niet enkel op instelling, maar ook op bron, thema, geldigheidsdatum of alfabetische volgorde opgezocht worden. Daarnaast is het mogelijk om de verschillende niveaus van documentatie (bestanden, variabelen, codes, enz) op een directe en interactieve manier te consulteren.
Alle vermelde informatie die momenteel beschikbaar is via de website van de KSZ is geïntegreerd in het CMS en zal via deze toepassing consulteerbaar zijn.

T2a: opstellen van documentatie bij de gegevens van de FOD SZ inzake personen met een handicap
De fiches horende bij de data van de FOD SZ werden opgesteld, gevalideerd door de FOD SZ en gepubliceerd op de website van de KSZ. Deze fiches hebben reeds een aantal revisies ondergaan.

T2b: opstellen van documentatie bij de gegevens van RSZ inzake de lokale vestigingen
De fiches RSZ werden aangevuld met gegevens uit StatbaseD die vanaf 2007 beschikbaar zijn in het DWH AM&SB. StatbaseD bevat sindsdien informatie over de lokale vestigingseenheid waar de werknemers tewerkgesteld zijn. Na validering door RSZ werden de fiches gepubliceerd op de website van de KSZ.

T2c: opstellen van documentatie bij de gegevens van de POD MI inzake activeringsmaatregelen in het kader van ESF
De fiches POD MI werden aangevuld met de beschrijving van de gegevens inzake activeringsmaatregelen in het kader van ESF. Na validering door de POD MI werden de fiches gepubliceerd op de website van de KSZ.

T2d: opstellen van documentatie bij de gegevens van andere eventueel nieuwe primaire en secundaire databanken
Het loopbaanbestand - werknemers dat in het verleden werd aangeleverd door CIMIRe wordt vanaf 2010 aangeleverd door Sigedis. Hierbij werden de loopbaangegevens van de periode voor 1990 geïntegreerd in het DWH AM&SB. De documentatie van deze gegevens werd opgesteld, gevalideerd door Sigedis en gepubliceerd op de website van de KSZ. Deze fiches hebben reeds een aantal revisies ondergaan.
Daarnaast werd er documentatie opgestelde bij de nieuwe gegevens in het DWH AM&SB. Het betreft de herkomstgegevens, de inkomensvariabelen en de variabelen inzake de LWI-indicator.
Met de Rijksdienst voor Jaarlijkse Vakantie vond overleg plaats i.v.m. de integratie van hun data in het DWH AM&SB. Hierbij werd een inventaris opgesteld met de beschikbare en bruikbare gegevens. De KSZ heeft een formele vraag voor integratie van de gegevens ingediend bij de RJV.
Ten slotte werd in kaart gebracht bij welke instelling gegevens inzake uitgaande grensarbeid beschikbaar zijn. Vermits een aanduiding van uitgaande grensarbeid beschikbaar is bij het NIC heeft de KSZ een formele vraag voor integratie van deze gegevens ingediend bij hen. Deze gegevens zijn reeds beschikbaar in het DWH AM&SB.

T3a: uitvoeren van een analyse die de inhoudelijke vergelijkbaarheid van het DWH AM&SB met de EAK (socio-economische positie volgens ILO-criteria) in kaart brengt

In een nota werd nagegaan in welke mate indicatoren voor nationale en internationale rapportering eveneens op basis van het DWH AM&SB kunnen worden berekend. Als beleidsdomein werd de EU 2020-strategie gekozen, meer bepaald de Work Intensity at household level. EU-SILC (en niet EAK) werd hiervoor als referentie genomen, zoals gebruikelijk voor Work Intensity op Europees niveau. De regels voor de constructie van deze indicator werden gepresenteerd aan de Gebruikersgroep. In overleg met ADSEI en betrokken dataleveranciers (RSZ, RSZPPO en RSVZ) werd de constructie verfijnd. Vervolgens werd een gegevensaanvraag ingediend om de LWI-indicator te testen. Ten slotte werd de documentatie opgesteld zodat de gegevens beschikbaar zijn voor de gebruikers van het DWH AM&SB. Deze documentatie omvat de beschrijving van de variabelen en een methodologische nota over de constructie van de variabelen. Daarnaast worden de eerste resultaten besproken in een bijdrage in het boek 'Het Datawarehouse, een duizendpoot! Perspectieven van het Datawarehouse Arbeidsmarkt en Sociale Bescherming - Le Datawarehouse, un mille-pattes! Perspectives du Datawarehouse Marché du Travail et Protection Sociale' (cfr. infra).

T3b: uitvoeren van een analyse die de inhoudelijke vergelijkbaarheid van het DWH AM&SB met de EU-SILC (inkomens- en armoedepositie volgens EU-SILC-criteria) in kaart brengt

Een vergelijking tussen de inkomensnotie van het DWH AM&SB en het inkomensconcept op basis van EU-SILC wordt beschreven in een bijdrage in het boek 'Het Datawarehouse, een duizendpoot! Perspectieven van het Datawarehouse Arbeidsmarkt en Sociale Bescherming - Le Datawarehouse, un mille-pattes! Perspectives du Datawarehouse Marché du Travail et Protection Sociale' (cfr. infra).

T4a: opstellen van een framework voor het inventariseren van de belangrijkste beleidsmaatregelen die impact hebben op de domeinen, de variabelen en codes van het DWH AM&SB

Er werd een framework opgesteld voor de inventarisatie van de belangrijkste beleidsmaatregelen die impact hebben op de populaties, de bronnen, de variabelen en de codes van het DWH AM&SB. Dit framework werd voorgesteld aan de Gebruikersgroep en aangepast aan de feedback van de gebruikers.

T4b: in een testfase invullen van het framework voor de RSZ, RVA en RVP voor de jaren 2003-2009

Het opgestelde framework werd ingevuld voor RSZ, RVA, FAO en RVP voor de periode 2005-2010. In de testfase werd het framework aangepast aan de bevindingen van de onderzoekers en aan de feedback van de Gebruikersgroep en van de betrokken instellingen. De juridische fiches van FAO, RVA, RSZ en RVP werden gevalideerd door de betrokken instellingen.
Wegens de beschikbaarheid van de bronnen en bruikbaarheid voor de gebruiker werd geopteerd om deze juridische fiche op te stellen voor de periode 2005-2010, in plaats van 2003-2009.
Na de evaluatie van deze testfase in de Gebruikersgroep werd beslist dat het niet zinvol en opportuun is om deze fiches uit te werken voor de overige bronnen. De juridische fiches van RVA, RVP en RSZ zijn echter wel beschikbaar voor de gebruikers via de website van de KSZ.

T5a: ontwikkelen van nieuwe afgeleide variabelen inzake herkomst

Tijdens een thematische Gebruikersgroep met betrekking tot het onderzoek naar de socio-economische positie van personen met vreemde origine werden de gebruikers geïnformeerd en geconsulteerd. Na overleg in de Gebruikersgroep werd op initiatief van de KSZ een procedure opgestart voor structurele integratie van afstammingsgegevens uit het Rijksregister in het DWH AM&SB. De KSZ diende bij het Rijksregister een aanvraag voor het verkrijgen van de gegevens. De eerste aanvraag bleek niet voldoende zodat de KSZ een aanvraag tot machtiging heeft aangevraagd bij het Sectoraal Comité van het Rijksregister. De machtiging van het Sectoraal Comité van het Rijksregister werd verleend en de data inzake herkomst werden geïntegreerd in

het DWH AM&SB (periode 2008-2012, die wordt gevolgd door een evaluatie). De nieuwe variabelen worden beschreven in fiches die beschikbaar zijn via de website van de KSZ.
Daarnaast werden er afgeleide variabelen inzake migratiehistoriek en nationaliteitshistoriek ontwikkeld. Deze variabelen werden besproken in de Gebruikersgroep en zijn reeds opgenomen in de webtoepassing globale cijfers.

T5b: ontwikkelen van nieuwe afgeleide variabelen inzake loopbaanperiodes
In overleg met Sigedis en de RVP werden er vijf afgeleide variabelen inzake loopbaanperiodes bepaald (gewerkte dagen, gewerkte dagen omgezet in VTE, gelijkgestelde dagen, gelijkgestelde dagen omgezet in VTE en bezoldiging). Deze variabelen werden besproken in de Gebruikersgroep. Vervolgens werden deze variabelen geconstrueerd.
Inzake de loopbaangegevens van zelfstandigen werd een eerste verkennende vergadering gehouden met het RSVZ. Hierbij werden de beschikbare gegevens bij het RSVZ (o.a. eClipz databank) geïnventariseerd.

T5c: ontwikkelen van nieuwe afgeleide variabelen inzake inkomen
Er werd een inkomensnotie (bruto en bruto belastbaar) uitgewerkt op basis van het DWH AM&SB. Deze notie bevat alle inkomens, zowel uit arbeid als uit uitkeringen, die beschikbaar zijn in het DWH AM&SB. Het inkomen werd geconstrueerd per aanleverende instelling (13) en dat op jaarbasis per individu. Deze constructie werd na validering door de instellingen gepresenteerd aan de Gebruikersgroep. Deze inkomensnotie werd tevens getest en de resultaten hiervan werden voorgelegd aan de betrokken instellingen. De variabelen worden beschreven in nieuwe fiches en de constructie ervan in een methodologische nota. Deze documentatie is beschikbaar via de website van de KSZ.
Daarnaast werd een inkomensnotie uitgewerkt die een 'beschikbaar inkomen' op basis van administratieve data beschrijft. Deze notie bevat een combinatie van inkomens beschikbaar in het DWH AM&SB (niet belastbare inkomens) en inkomens beschikbaar bij de FOD Financiën – IPCAL (belastbare inkomens). Deze nota werd in samenspraak met de FOD Financiën opgesteld.

T6a: uitwerken en ontwikkelen van webtoepassingen inzake socio-economische mobiliteit
In overleg met Smals, en na de consulatie van de Gebruikersgroep, werd de webtoepassing socio-economische mobiliteit ontwikkeld (2012). Deze webtoepassing maakt het mogelijk om de socio-economische evolutie van de populatie opgenomen in het DWH AM&SB (op lange termijn, met name per jaar) op te volgen doorheen de tijd, zowel vooruit- als achteruitkijkend. Volgende parameters zijn opgenomen in de toepassing: geslacht, leeftijd, gewest, nomenclatuur, aanduiding overlijden en arbeidsregime. Deze webtoepassing is beschikbaar via de website van de KSZ.

T6b: uitwerken en ontwikkelen van een webtoepassing inzake gezinssamenstelling
De webtoepassing inzake gezinssamenstelling werd inhoudelijk uitgewerkt en besproken in de Gebruikersgroep. Tevens werd een behoefteanalyse opgesteld door Smals. Echter, wegens budgettaire redenen is deze webtoepassing nog niet beschikbaar.

T7: opvolgen van de nomenclatuur van de socio-economische positie
De nomenclatuur van de socio-economische positie werd permanent geactualiseerd. Daarnaast werden volgende aanpassingen aangebracht:

- De benaming van positie n371 is aangepast (nieuwe benaming: 'arbeidsongeschiktheid gekend bij de mutualiteiten').
- Nieuwe afgeleide variabelen horende bij de nomenclatuur werden toegevoegd.
- Nieuwe waarden van de variabelen van de RVA werden geïntegreerd in de nomenclatuur.
- De nomenclatuur werd aangepast aan wijzigingen in het Pensioenkadaster (RVP).
- De populatie van het FAO werd geïntegreerd in de nomenclatuur (n374

'arbeidsongeschikt omwille van een arbeidsongeval') en er werden afgeleide variabelen voor deze populatie geconstrueerd.
- De populatie van de FOD SZ werd geïntegreerd in de nomenclatuur (n38 'persoon met tegemoetkoming aan personen met een handicap'). Tevens werden er nieuwe afgeleide variabelen voor deze populatie gecreëerd.
- De terbeschikkingstelling voorafgaand aan het pensioen werd (vanaf 2011) geïdentificeerd in de nomenclatuur en uitgedrukt in een nieuwe positie (n352 'terbeschikkingstelling voorafgaand aan het pensioen') en bijhorende afgeleide variabelen.

Door de integratie van nieuwe gegevens in het DWH AM&SB en in de nomenclatuur verkleinde de groep van personen die niet gekend zijn met een nomenclatuurpositie (n4).

T8: inzetten van de expertise inzake de effectieve koppeling van diverse Belgische statistische bronnen aan het DWH AM&SB door middel van samenwerking met andere overheden
De onderzoeksploeg diende twee gegevensaanvragen in die een koppeling van het DWH AM&SB met gegevens van andere overheden beogen. De eerste aanvraag betreft een koppeling met de EU-SILC (ADSEI), de tweede aanvraag betreft een koppeling met de IPCAL-gegevens van de FOD Financiën. Beide aanvragen zijn momenteel nog niet uitgevoerd en konden bijgevolg niet worden geanalyseerd.

T9a: geven van feedback in de jaarlijkse Gebruikersgroep
Zie sectie 'Gebruikersgroep'.

T9b: geven van feedback in de werkgroep inzake herkomstgegevens
Zie T5a.

T9c: geven van feedback in de werkgroep inzake de uitwerking van de inkomensnotie op basis van het DWH AM&SB
Na de bespreking van de inkomensnotie in de Gebruikersgroep werd een werkgroep opgericht. Deze werkgroep staat momenteel in zijn kinderschoenen maar heeft de verdere uitwerking van een 'beschikbaar inkomen' op basis van administratieve data tot doel gesteld. Hierbij zal onder andere gebruik worden gemaakt van de IPCAL-gegevens van de FOD Financiën.

T10a: bieden van methodologische ondersteuning aan de werkgroep 'omnisectoriële impact SZ-uitkeringen'
Er werd antwoord gegeven op de vragen van de de werkgroep 'omnisectoriële impact SZ-uitkeringen'. Tevens werden er een aantal variabelen geconstrueerd in het kader van de 'omnisectoriële impact SZ-uitkeringen'. Deze informatie werd opgeleverd ten aanzien van de FOD SZ in twee documenten, één m.b.t. de gezinsstatuten en één m.b.t. de socio-economische statuten.

T11: organiseren van een internationaal expertseminarie
Het internationaal seminarie van het project DOCDWH vond plaats op 15 september 2010 in de Faculty Club, Leuven. Twee buitenlandse experten lichtten databanken opgebouwd aan de hand van administratieve data toe (prof. Bart Bakker: Het Sociaal Statistisch bestand, CBS Nederland en Pernille Stender: RAS-Statistics, Statistics Denmark).

Tussentijds overleg

Maandelijks kwamen de onderzoeksteams samen met de KSZ en FOD SZ om de stand van zaken van de uit te voeren taken te overlopen en de werkzaamheden op te volgen. De

verslagen van deze maandelijkse vergaderingen kunnen worden opgevraagd bij de coördinator van het project.

Begeleidingscomité

Bij aanvang van het project werd in samenspraak met Federaal Wetenschapsbeleid, de KSZ en de FOD SZ beslist om het begeleidingscomité van het project DOCDWH te laten samenvallen met de reeds bestaande Gebruikersgroep van het DWH AM&SB. Deze Gebruikersgroep komt ad hoc samen, met een minimum van één maal per jaar. De leden van de Gebruikersgroep maken allen gebruik van het DWH AM&SB en het betreffen voornamelijk (academische) onderzoekers, beleidsmakers en dataleveranciers.

Gedurende de looptijd van het project kwam de Gebruikersgroep zeven keer samen. Naast een stand van zaken i.v.m. het DWH AM&SB (bronnen, basis- en webtoepassingen, documentatie) kwamen volgende topics aan bod:

- 28 juni 2010: De integratie van de gegevens inzake herkomst in het DWH AM&SB stond centraal.
- 15 september 2010: De tweede bijeenkomst van de Gebruikersgroep viel samen met het internationaal seminarie ("Databases compiled on the basis of administrative registers"). Hierbij werden twee internationale experten uitgenodigd (prof. Bart Bakker en Pernille Stender).
- 20 december 2010: Tijdens deze Gebruikersgroep lag de nadruk op de voorstelling van de inkomensnotie op basis van het DWH AM&SB.
- 23 november 2011: De LWI-indicator op basis van het DWH AM&SB en een analyse van gezondheidsberoepen kwamen aan bod in deze Gebruikersgroep.
- 28 juni 2012: De eerste resultaten van de LWI-indicator op basis van het DWH AM&SB werden voorgesteld tijdens deze Gebruikersgroep. Daarnaast werd een stand van zaken gegeven i.v.m. de inkomensnotie en de indicatoren over de levenscyclus en de arbeidsmarkt.
- 26 september 2012: De eerste resultaten op basis van de herkomstgegevens uit het DWH AM&SB, een voorstel i.v.m. herkomstvariabelen en een classificatie voor de opname van de herkomstgegevens in de toepassingen van het DWH AM&SB werden besproken tijdens deze Gebruikersgroep.
- 26 juni 2013: Deze Gebruikersgroep gaf een stand van zaken i.v.m. de nieuwe webtoepassing inzake gezinssamenstelling, de fiches omtrent primaire databanken, de juridische fiches, de loopbaancodes van het loopbaanbestand (loontrekkenden) en de biografische indicatoren.

Boek inzake het DWH AM&SB

Naar aanleiding van het einde van het project DOCDWH was het idee gerezen om een boek te publiceren i.v.m. het DWH AM&SB. Het boek geeft een overzicht van de mogelijkheden van het DWH AM&SB. Ten eerste bevat het boek een (beknopt) historisch overzicht en een stand van zaken van het DWH AM&SB. Ten tweede komen enkele onderzoeken op basis van het DWH AM&SB aan bod. De beschrijving van deze onderzoeken schetst een goed beeld van de mogelijkheden van het DWH AM&SB. Er wordt een zo breed mogelijke waaier aan onderzoeken beschreven zodat het DHW AM&SB in al zijn facetten aan bod komt. Vervolgens worden enkele toepassingen die gebruik maken van data uit het DWH AM&SB toegelicht. Ten slotte worden indicatoren (biografische gebeurtenissen, LWI-indicator en de inkomensnotie) die kunnen worden berekend op basis van het DWH AM&SB, besproken. Er is niet geopteerd om alle hoofdstukken zelf te schrijven maar wel om met partners te werken.

Het boek kreeg de titel 'Het Datawarehouse, een duizendpoot! Perspectieven van het Datawarehouse Arbeidsmarkt en Sociale Bescherming - Le Datawarehouse, un mille-pattes! Perspectives du Datawarehouse Marché du Travail et Protection Sociale' (editors Hans Knapen, Stijn Braes, Thomas Ermans en Wim Herremans). Dit boek wordt uitgegeven door Academia Press in de reeks 'Wetenschap en Maatschappij'. Deze reeks publiceert eindrapporten van sociaal-wetenschappelijke onderzoeksprojecten die uitgeschreven en gefinancierd zijn door Federaal Wetenschapsbeleid, onder andere in het kader van het Agora-programma. Het boek geldt bijgevolg mede als eindverslag van het project DOCDWH. Het boek is verschenen in februari 2014.

Lien vers le site web du projet

- http://www.ksz.fgov.be/nl/bcss/nodepage/content/websites/belgium/statistics/statistics_01.html

- FRANCAIS -

Introduction

Le projet DOCDWH comprend l'approfondissement du contenu et des développements méthodologiques du Datawarehouse Marché du Travail et Protection Sociale (DWH MT&PS). Ceci implique (1) l'approfondissement de l'offre initiale en données et métadonnées et (2) divers compléments qui y sont apportés.

Le projet DOCDWH a été mené sous la direction du Prof J. Berghman (Centrum voor Sociologisch Onderzoek – KU Leuven), opérant en tant que coordinateur, et des promoteurs Prof. P. Desmarez (METICES/TEF – Université Libre de Bruxelles) et Prof. L. Sels (Steunpunt WSE – KU Leuven). Le projet, exécuté à la demande de la Banque Carrefour de la Sécurité Sociale (BCSS) et du SPF Sécurité Sociale (SPF SS), s'est déroulé de décembre 2009 à février 2014.

Contexte

La réalisation des deux objectifs principaux s'est appuyée sur le projet AGORA DW-MTPS (AG/01/115), dont le but était de mener une analyse approfondie du contenu des sources figurant dans le DWH MT&PS et à rédiger un manuel méthodologique. Ce projet visait l'exploitation des sources de données en vue de l'amélioration du rapportage quantitatif, l'analyse de propositions politiques et la réalisation de recherches liées à la protection sociale, par le SPF SS ou d'autres. Au terme du projet en 2009, la documentation créée a été mise à la disposition du public via le site Internet de la BCSS, sous forme de fichiers téléchargeables. Cette documentation représente l'offre de base pour les utilisateurs.

Cependant, le DWH MT&PS n'est pas un ensemble statique. Les modifications apportées à la législation, les adaptations de la réglementation, les réorganisations au sein de la sécurité sociale, les réformes techniques des bases de données, les changements de variable, etc. ont un impact sur les bases de données primaires et secondaires sur lesquelles repose le DWH MT&PS. Ces événements peuvent également avoir un impact direct sur le DWH MT&PS. De plus, le nombre de sources de données s'élargit continuellement. Tous ces facteurs ont de ce fait un impact important sur la documentation et doivent être pris en compte en permanence par des mises à jour.

Objectifs et tâches accomplis

Les deux objectifs centraux ont été déclinés en plusieurs sous-objectifs, soit:

(1) Les possibilités offertes par le DWH MT&PS pour l'analyse socio-économique et le rapportage national et international doivent être examinées. L'attention doit être portée sur la comparaison avec d'autres données de recherche socio-économique, les possibilités et lacunes pour l'interprétation des données ainsi que les forces et faiblesses du DWH MT&PS.

(2) De nouvelles notions doivent être élaborées. Il s'agit ici de nouvelles variables dérivées. Ces variables démontrent la valeur ajoutée du DWH MT&PS, c'est-à-dire le

couplage de données issues de sources différentes.

(3) L'impact sur le DWH MT&PS de la législation et des modifications qui y sont apportées doit être estimé plus finement. A cet effet, un inventaire doit être établi des principales mesures législatives qui ont un impact sur les populations, les sources, les variables et les codes du DWH MT&PS. Enfin, il faut évaluer dans quelle mesure l'impact de telles modifications peut être apprécié à l'aide du DWH MT&PS.
(4) La manière avec laquelle les données sont reprises dans les bases de données primaires doit être davantage documentée. Il s'agit d'un travail essentiel afin d'obtenir une image claire de la qualité des données.
(5) Une actualisation de la documentation fournie dans le cadre du projet DW-MTPS ainsi que de l'intégration des données 'VDAB/FOREM/Actiris/ADG' (projet AGORA AG/00/143) doit être prévue. Enfin, l'entièreté de la documentation doit être intégrée dans un système de documentation des métadonnées.
(6) Il faut examiner dans quelle mesure les autorités des différents départements peuvent collaborer à la diffusion de données dans le cadre de projets de recherche socio-économique dont le domaine dépasse celui de leur département.

Afin de réaliser ces objectifs, le travail a été réparti en 11 tâches et sous-tâches. Celles-ci sont présentées dans la section ci-dessous avec les résultats.

Résultats

T1a: Actualiser la documentation et le manuel qui a été fourni dans le cadre des projets AG/01/115 et AG/00/143

Les fiches ONSS, ONSSAPL, INASTI, VDAB-Forem-Actiris-ADG, SdPSP, CIMIRe/Sigedis, ONP, ONEm, SPP IS, SPP IS FSE, FMP, SPF SS, ONAFTS, INASTI – allocations familiales, FAT, Registre National, minirn, INASTI-carrières, données origines, CIN, nomenclature socio-économique, manuel d'utilisateur, revenus et la description des conditions pour l'établissement des positions de la nomenclature ont été actualisées de manière permanente. Le fichier des carrières – salariés, qui par le passé était fourni par CIMIRe, est depuis 2010 fourni par Sigedis. Une nouvelle documentation relative à cette source a été créée.

T1b : Ajout aux métadonnées des informations sur la manière dont les données sont reprises dans les bases de données primaires

Afin de mieux estimer la qualité des données du DWH MT&PS, il est nécessaire de représenter la manière dont celles-ci sont initialement collectées. Dans une première phase, un 'template' (fiche) a été créé pour pouvoir documenter de manière uniforme. Ensuite, en collaboration avec les institutions concernées, les fiches ont été remplies. L'information est aujourd'hui disponible sur le site Internet de la BCSS.

T1c : Intégration de la documentation existante et nouvelle dans le Content Management System qui devra encore être développé

Premièrement, en collaboration avec Smals, la documentation, existante et nouvelle, a été intégrée sur Beconnected (une application Alfresco) en 2011. Cette plateforme en ligne intègre à un workflow la procédure par laquelle des modifications sont apportées à la documentation (modification du contenu – traduction – publication). L'actualisation et la modification de toute la documentation se déroule depuis via cette plateforme. Les chercheurs y ont accès via un lecteur de carte eID.

De plus, préalablement à son intégration dans le Content Management System (CMS), la documentation a été modifiée, aussi bien du point de vue de la forme que du contenu (dénomination des fichiers, numérotation des variables et des codes, uniformisation des tables, etc.). La réalisation fonctionnelle du CMS a déjà été entamée durant le précédent projet DW-MTPS mais a pris un retard important durant le projet DOCDWH. Aujourd'hui, toute la documentation

est cependant bien chargée dans le système. L'équipe des chercheurs a testé le CMS de long en large, que ce soit au niveau de la cohérence, de la complétude ou de la convivialité. Lorsque les derniers 'bugs' auront été résolus par la Smals, le CMS sera présenté au groupe d'utilisateurs (dont la réunion est prévue pour le deuxième trimestre de 2014). Ce système remplace les documents Word qui étaient proposés via le site Internet de la BCSS. Le CMS rend la consultation de toute la documentation relative au DWH MT&PS plus conviviale. Les variables peuvent ainsi être recherchées non seulement par institution mais également selon la source, le thème, la date de validité ou l'ordre alphabétique. De plus, différents niveaux de la documentation (fiches, variables, codes, etc.) peuvent être consultés de manière plus directe et interactive.
Toute l'information communiquée actuellement via le site de la BCSS est intégrée au CMS et sera consultable via cette application.

T2a : Constituer la documentation relative aux données du SPF Sécurité Sociale concernant les personnes souffrant d'un handicap
Les fiches ayant trait aux données du SPF SS ont été créées, validées par le SPF SS et publiées sur le site de la BCSS. Ces fiches ont déjà connu un certain nombre de révisions.

T2b : Constituer la documentation des données de l'ONSS concernant les établissements locaux
Les fiches ONSS ont été complétées des données de StatbaseD, une source disponible depuis 2007 dans le DWH MT&PS. StatbaseD contient l'information sur les établissements locaux où sont employés les salariés. Après validation par l'ONSS, les fiches ont été publiées sur le site Internet de la BCSS.

T2c : Constituer la documentation des données du SPP IS relatives aux mesures d'activation dans le cadre du FSE
Les fiches SPP IS ont été complétées de la description des données relatives aux mesures d'activation dans le cadre du FSE. Après validation par le SPP IS, les fiches ont été publiées sur le site Internet de la BCSS.

T2d : Constituer la documentation d'éventuelle nouvelles bases de données primaires ou secondaires
Le fichier 'carrière – salariés' qui était auparavant fourni par CIMIRe est dépuis 2010 fourni par Sigedis. Les données de carrière pour la période précédent 1990 ont ici pu être intégrées au DWH MT&PS. La documentation de ces données a été créée et validée par Sigedis, puis publiée sur le site Internet de la BCSS. Ces fiches ont déjà fait l'objet de plusieurs révisions.
Une nouvelle documentation a également du être établie pour les nouvelles données du DWH MT&PS. Il s'agit ici des données d'origine, des variables sur les revenus et de variables relatives à l'indicateur LWI.
Une concertation a eu lieu avec l'Office National des Vacances Annuelles afin d'intégrer leur données au DWH MT&PS. De ce point de vue, un inventaire des données disponibles et utilisables a été constitué. La BCSS a introduit une demande formelle à l'ONVA pour l'intégration de leurs données.
Finalement, il a été établi quelles institutions disposaient de données relatives au travail frontalier. Vu que l'indication de travail frontalier est disponible au CIN, la BCSS y a introduit une demande formelle d'intégration de données. Ces données sont désormais disponibles dans le DWH MT&PS.

T3a : Effectuer une analyse qui mette en lumière la comparabilité du contenu du DWH avec l'EFT (position socioéconomique selon l'OIT)
Dans une note, il a été évalué dans quelle mesure les indicateurs pour le rapportage national et international pouvaient être calculés sur base du DWH MT&PS. Le cadre politique choisi a été la stratégie EU-2020, et plus précisément l'intensité de travail - Work Intensity – au niveau du ménage. L'enquête EU-SILC (et pas l'EFT) a été ici prise comme référence, puisque c'est sur celle-ci que repose l'indicateur Work Intensity au niveau européen. Les règles de construction de ces

indicateurs ont été présentées au groupe d'utilisateurs. En collaboration avec la DGSIE et les fournisseurs de données concernés (ONSS, ONSSAPL et INASTI), la construction de l'indicateur a été affinée. Ensuite, une demande de données a été introduite auprès de la BCSS afin de tester l'indicateur LWI. Finalement, une documentation a été constituée afin de rendre accessible l'indicateur aux utilisateurs. Celle-ci contient la description des variables et une note méthodologique sur la construction des variables. De plus, les premiers résultats ont été discutés dans une contribution au livre 'Le datawarehourse, un mille-pattes !' Perspectives du Datawarehouse marché du travail et Protection sociale' (cfr. Infra).

T3b : Effectuer une analyse qui mette en lumière la comparabilité du contenu du DWH avec l'EU-SILC (position en matière de revenus et de pauvreté selon les critères EU-SILC)
Une comparaison entre la notion de revenus du DWH MT&PS et du concept de revenus sur base d'EU-SILC est décrite dans une contribution au livre 'Le datawarehourse, un mille-pattes !' Perspectives du Datawarehouse marché du travail et Protection sociale' (cfr. Infra).

T4a : Constituer un schéma pour l'inventorisation des principales mesures stratégiques qui ont un impact sur les domaines, les variables et les codes du DWH MT&PS
Un schéma a été constitué pour l'inventorisation des principales mesures stratégiques qui ont un impact sur les domaines, les variables et les codes du DWH MT&PS. Ce schéma a été présenté au groupe d'utilisateurs et mis à jour en accord avec le retour des utilisateurs.

T4b : Dans une phase de test, compléter le schéma pour l'ONSS, l'ONEm et l'ONP pour les années 2003-2009
Le schéma mis au point a été complété pour l'ONSS, l'ONEm, le FAT et l'ONP pour la période 2005-2010. Dans la phase de test, le schéma a été modifié pour tenir compte des remarques des chercheurs et du retour du groupe d'utilisateurs ainsi que des institutions concernées.
Dans le cas où les sources étaient disponibles et l'utilité pour l'utilisateur avérée, des fiches ont été constituées pour la période 2005-2010 plutôt que 2003-2009.
Après évaluation de cette phase de test au sein du groupe d'utilisateur, il a été décidé que la réalisation de ces fiches n'était pas opportune pour le reste des sources de données. Les fiches juridiques de l'ONSS, l'ONEm et l'ONP sont cependant bien disponibles aux utilisateurs via le site Internet de la BCSS.

T5a : Développer de nouvelles variables dérivées concernant l'origine
Durant une réunion thématique du groupe d'utilisateurs relative à la recherche sur base de la position socio-économique des personnes d'origine étrangère, les utilisateurs ont été informés et consultés. Après concertation au sein du groupe d'utilisateur, la BCSS a initié une procédure d'intégration structurelle des données d'origine issues du Registre National, dans le DWH MT&PS. Une première demande formelle a ainsi été introduite, jugée non-satisfaisante en l'absence d'une demande préalable d'autorisation au Comité Sectoriel du Registre national. L'autorisation du Comité Sectoriel du registre national a finalement été donnée et les données d'origine ont ainsi été intégrées dans le DWH MT&PS (pour la période 2008-2012, suivie d'une évaluation). Les nouvelles variables sont décrites dans des fiches disponibles via le site Internet de la BCSS.
De plus, des variables dérivées relatives à l'historique des migrations et à l'historique des nationalités ont été développées. Ces variables ont été discutées dans une réunion du groupe d'utilisateurs et sont déjà reprises dans les applications web 'Chiffres globaux'.

T5b : Développer de nouvelles variables relatives aux périodes de carrière
En concertation avec Sigedis et l'ONP, cinq variables relatives aux périodes de carrière ont été définies (jours prestés, jours prestés convertis en ETP, jours assimilés, jours assimilés convertis en ETP et traitements). Ces variables ont été discutées au sein du groupe d'utilisateurs. Elles ont ensuite été construites.
En ce qui concerne les carrières des travailleurs indépendants, une première réunion, pour nouer contact, a eu lieu avec l'INASTI. A cette occasion, les données disponibles à l'INASTI (provenant

entre autres de la base de données eClipz) ont été inventoriées.

T5c : développer de nouvelles variables dérivées relatives aux revenus
Une notion de revenu (brut et brut imposable) a été implémentée sur base du DWH MT&PS. Cette notion comprend tous les revenus, aussi bien du travail que des allocations, qui sont disponibles au sein du DWH MT&PS. Le revenu a été reconstruit par institution fournisseuse (13), sur base annuelle et par individu. Après validation par les institutions concernées, la notion a été présentée au groupe d'utilisateurs. Elle a également été testée et les résultats des tests présentés aux institutions concernées. Les variables sont décrites dans de nouvelles fiches et leur construction dans une note méthodologique. Cette documentation est disponible via le site Internet de la BCSS.
En outre, une notion de revenus a été développée qui décrit un 'revenu disponible' sur base de données administratives. Cette notion repose sur une combinaison des revenus disponibles dans le DWH MT&PS (revenus non imposables) et les revenus disponibles auprès du SPF Finances – IPCAL (revenus imposables). Cette note a été établie en collaboration avec le SPF Finances.

T6a : Evaluer et développer des applications web concernant la mobilité socio-économique
En concertation avec Smals et après consultation du Groupe d'utilisateurs, une application mobilité socio-économique a été développée (2012). Cette application permet de suivre l'évolution socio-économique de populations reprises dans le DWH MT&PS (à long terme, c'est-à-dire par année), aussi bien de manière prospective que rétrospective. Les paramètres suivants sont repris dans l'application : sexe, âge, région, position sur la nomenclature socio-économique, indication de décès et régime de travail. Cette application est disponible via le site Internet de la BCSS.

T6b : Evaluer et développer des applications web concernant la composition des ménages
Le contenu de l'application relative à la composition des ménages a été développé et discuté au sein du groupe d'utilisateur. Une analyse des besoins a également été réalisée par Smals. Cependant, en raisons de contraintes budgétaires, l'application n'est pas encore disponible.

T7 : Assurer un suivi de la nomenclature socio-économique
La nomenclature des positions socio-économiques a été mise à jour de manière continue. En outre, les modifications suivantes y ont été apportées :

- L'appellation de la position n371 a été modifiée (nouvelle appellation : 'incapacité de travail connue des mutuelles').
- De nouvelles variables dérivées ont été ajoutées à la nomenclature.
- De nouvelles valeurs et variables de l'ONEm ont été intégrées dans la nomenclature.
- La nomenclature a été adaptée aux modifications du cadastre des pensions (ONP).
- La population du FAT a été intégré dans la nomenclature (n374 'incapacité de travail en raison d'un accident de travail') et des variables dérivées relatives à cette population ont été créées.
- La population du SPF SS a été intégrée dans la nomenclature (n38 'Personne avec une allocation aux personnes handicapées') et des variables dérivées relatives à cette population ont été créées.
- La mise à disposition avant la pension a été identifiée dans la nomenclature (depuis 2011) et exprimée dans une nouvelle position (n352 'Mise à disposition avant la pension') ainsi que des variables dérivées.

Par l'intégration de nouvelles données dans le DWH MT&PS, le groupe de personnes non connues (position 'n4') a été amoindri.

T8 : Engager l'expertise nécessaire quant à la liaison effective de diverses sources statistiques belges avec le DWH MT&PS par le biais d'une collaboration avec d'autres institutions publiques
L'équipe des chercheurs a introduit deux demandes de données qui visaient un couplage de

données avec des institutions publiques externes au DWH MT&PS. La première demande a trait à un couplage avec EU-SILC (DGSIE) et la seconde à un couplage avec les données IPCAL du SPF Finances. Ces deux couplages n'ont pour l'instant pas encore exécuté et leurs objets ne peuvent donc pas être analysés.

T9a : Donner un feedback au sein du groupe annuel d'utilisateurs
Voir section 'groupe d'utilisateurs'.

T9b : Donner un feedback au sein du groupe de travail relatif aux données d'origine
Voir T5a.

T9c : Donner un feedback au sein du groupe de travail relatif à la notion de revenus sur base du DWH MT&PS
Après discussion de la notion de revenus au sein du groupe d'utilisateurs, un groupe de travail a été mis sur pied. Celui-ci n'en est pour le moment qu'à ses balbutiements mais s'est fixé comme objectif l'approfondissement de la notion de 'revenu disponible' sur base de données administratives. Dans cette optique, il sera entre autres fait usage des données IPCAL du SPF Finances.
T10a : Offrir un support méthodologique au groupe de travail 'impacts omnisectoriels allocations de sécurité sociale'
Des réponses ont été apportées aux questions du groupe de travail 'impacts omnisectoriels allocations de sécurité sociale'. Un certain nombre de variables ont également été construites dans le cadre de 'impacts omnisectoriels allocations de sécurité sociale'. Cette information a été fournie à l'attention du SPF SS en deux documents, le premier relativement aux statuts des ménages et le second aux statuts socio-économiques.

T11 : Organiser un séminaire international d'experts
Le séminaire international du projet DOCDWH a eu lieu le 15 septembre 2010 dans le Faculty Club, Leuven. Deux experts internationaux y ont présenté des bases de données construites sur des données administratives (prof. Bart bakker : Het sociaal Statistisch bestand, CBS Nederland en Pernille Stender : RAS-Statistics, Statistics Denmark).

Réunions intermédiaires

Les équipes de chercheurs se réunissent mensuellement avec la BCSS et le SPF SS pour parcourir l'état des lieux des tâches à réaliser et suivre l'activité de chacun. Les rapports de ces réunions mensuelles peuvent être obtenus auprès du coordinateur du présent projet.

Comité d'accompagnement

Au commencement du projet, en concertation avec la politique scientifique fédérale, il a été décidé de faire coïncider le comité d'accompagnement du projet DOCDWH avec l'alors déjà existant groupe d'utilisateurs du DWH MT&PS. Ce groupe d'utilisateurs se réunit quand nécessaire, au minimum une fois par an. Les membres de ce groupe d'utilisateurs font tous usage du DWH MT&PS et se retrouvent essentiellement parmi les chercheurs (universitaires), les décideurs politiques et les fournisseurs de données.

Durant le cours du projet, le groupe d'utilisateurs s'est réuni sept fois. En plus d'un état des lieux du DWH MT&PS (sources, applications web/de base, documentation), les sujets suivants ont été abordés :

- 28 juin 2010 : L'intégration des données relatives à l'origine dans le DWH MT&PS a été le sujet central

- 15 septembre 2010 : La deuxième tenue du groupe d'utilisateurs a coïncidé avec le séminaire international ('Databases compiled on the basis of administrative registers'). Deux experts internationaux y étaient conviés (prof. Bart bakker : Het sociaal Statistisch bestand, CBS Nederland en Pernille Stender : RAS-Statistics, Statistics Denmark).
- 20 décembre 2010 : Durant ce groupe d'utilisateurs, l'attention fut portée à la notion de revenus sur base du DWH MT&PS.
- 23 novembre 2011 : Durant ce groupe d'utilisateurs, ont été abordés l'indicateur LWI ainsi qu'une analyse des métiers de la santé.
- 28 juin 2012 : Les premiers résultats de l'indicateur LWI y ont été présentés. Un état des lieux de la notion de revenus et des indicateurs cycles de vie a également été proposé.
- 26 septembre 2012 : Y ont été discuté, les premiers résultats des données d'origines récemment intégrées au DWH MT&PS ainsi qu'une proposition de classification des données d'origine à utiliser dans les applications du DWH MT&PS.
- 26 juin 2013 : Durant ce groupe d'utilisateurs, ont été présentés un état des lieux des nouvelles applications relatives à la composition des ménages, les fiches des bases de données primaires, les codes carrières du fichier carrière (salariés) ainsi que les indicateurs biographiques.

Livre sur le DWH MT&PS

L'idée de publier un livre sur le DWH MT&PS a été soulevée parmi les chercheurs à l'occasion de la fin du projet. Le livre offre une vue d'ensemble des possibilités du DWH MT&PS. Premièrement, le livre fournit un historique (succinct) et un état des lieux du DWH MT&PS. Deuxièmement, diverses études reposant sur l'analyse de données du DWH MT&PS sont présentées. La description de ces études dresse un portrait représentatif et un éventail aussi large que possible des possibilités offertes pour l'analyse scientifique. Certaines applications faisant usage du DWH MT&PS sont ensuite présentées. Finalement, divers indicateurs (indicateurs biographiques, indicateur LWI, notion de revenus) pouvant être calculés sur base du DWH MT&PS sont discutés. L'objectif de l'équipe des chercheurs n'était pas ici de rédiger tous les articles mais bien de travailler avec différents partenaires.

Le livre s'intitule 'Het Datawarehouse, een duizendpoot! Perspectieven van het Datawarehouse Arbeidsmarkt en Sociale Bescherming - Le Datawarehouse, un mille-pattes! Perspectives du Datawarehouse Marché du Travail et Protection Sociale' (editeurs Hans Knapen, Stijn Braes, Thomas Ermans en Wim Herremans). Il est publié par Academia Press dans la série 'Problèmes actuels concernant la cohésion sociale'. Cette série a pris l'habitude de publier les rapports de fin de projet dans le domaine des sciences sociales pour des équipes de recherche financées par la Politique scientifique fédérale, notamment dans le cadre du programme AGORA. Le livre a de ce fait aussi valeur de rapport final pour le projet DOCDWH. Le livre est paru en février 2014.

Links to the project website

- http://www.ksz.fgov.be/nl/bcss/nodepage/content/websites/belgium/statistics/statistics_01.html

- ENGLISH -

Introduction

The DOCDWH project covered the content expansion and methodological development of the Datawarehouse Labour Market and Social Protection (DWH LM&SP) whereby (1) the initial basic collection of the data originating from the social security institutions and the metadata was expanded and (2) supplemented with new data and metadata.

The DOCDWH project was led by Prof. J. Berghman (Centrum voor Sociologisch Onderzoek – KU Leuven), who served as coordinator, as well as promoters Prof. P. Desmarez (METICES/TEF – Université Libre de Bruxelles) and Prof. L. Sels (Steunpunt Werk en Sociale Economie – KU Leuven). The project, which was implemented at the request of the Crossroads Bank for Social Security (CBSS) and the Federal Public Service Social Security (FPS SS), ran from December 2009 to February 2014.

Context

The achievement of the two central objectives built upon the AGORA DW-AMSB project (AG/01/115), the goal of which was to conduct a thorough substantive analysis of the sources entered into the DWH LM&SP and to create a methodological handbook. This project was specifically aimed at improving access to data sources for the purpose of improving the quantitative reporting, the analysis of policy proposals and the conducting of research related to the social protection offered by the FPS SS or others. When this project was completed in 2009, its documentation was made available to the general public via the CBSS website in the form of downloadable documents. This documentation can be considered to be a basic package.
However, the DWH LM&SP is not a static entity. Changes in legislation, the reworking of regulations, reorganisations within social security, technical reforms to databases, changing variables, etc., all have an impact on the primary and secondary databases upon which the DWH LM&SP is built. These events can also have a direct influence on the DWH LM&SP, and the number of sources is moreover gradually increasing. All of these factors therefore have an impact on the documentation, and they must be compensated for on a permanent basis.

Objectives and implemented tasks

The two central objectives were divided into several sub-objectives, namely:

(1) The possibilities offered by the DWH LM&SP for socio-economic research and national and international reporting must be analysed. This must include a focus on the comparison with different sources for socio-economic research, the difficulties and potential pitfalls when interpreting the data and the strengths and weaknesses of the DWH LM&SP.

(2) A number of new concepts must be developed, namely new derived variables. These variables demonstrate the immediate added value of the DWH LM&SP, i.e. the linking of data from different sources.

(3) The impact of laws and of changes to these laws on the DWH LM&SP must be specified in further detail. To do this, an inventory must be drawn up of the most

significant measures having an impact on the populations, sources, variables and codes of the DWH LM&SP. In addition, the extent to which the impact of these measures can be measured by means of the DWH LM&SP must be determined.

(4) The way in which the data are entered into the primary databases must be further documented. This is critical in order to get a good sense of the quality of the data.

(5) The documentation that was delivered within the scope of the DW-AMSB project and the 'integration of data from the VDAB/FOREM/Actiris/ADG into the DWH LM&SP' project (AGORA project with reference AG/00/143) must be made available for the updating process. Furthermore, all of the documentation must be integrated into a metadata documentation system.

(6) The extent to which government agencies in different departments are able to cooperate in data-sharing within the context of socio-economic research projects that transcend the domain of a single government department must be determined.

In order to achieve the stated sub-objectives and results, the work was divided into eleven tasks and sub-tasks. These tasks/sub-tasks are described in the following section with the results.

Results

T1a: update the documentation and the handbook that were delivered as part of projects AG/01/115 and AG/00/143

The information that was updated on a permanent basis includes the dossiers from the National Social Security Office (NSSO), National Social Security Office for Provincial and Local Government (NSSOPLG), National Institute for the Social Security of the Self-Employed (NISSE), the different regional employment services in Belgium (VDAB-FOREM-Actiris-ADG), Pension Service for the Public Sector (PDOS), CIMIRe/Sigedis, National Pensions Office (RVP), National Employment Office (NEO), Programmatic Public Service Social Integration anti-Poverty Policy, Social Economy (PPS SI), Fund for Occupational Diseases (FBZ), Federal Public Service Social Security (FPS SS), National Child Benefit Service for Employees (RKW), National Institute for the Social Security of the Self-Employed - child benefits, Programmatic Public Service Social Integration anti-Poverty Policy, Social Economy - European Social Fund (PPS SI - ESF), Fund for Occupational Accidents (FAO), Belgian National Register, minirn database, National Institute for the Social Security of the Self-Employed - careers, origin data, National Intermutualist College (NIC), nomenclature, user guide, income and the description of the conditions for the nomenclature. The career database - employees, which in the past had been supplied by CIMIRe, has since 2010 been supplied by Sigedis. New documentation related to this source was created.

T1b: supplement the metadata with information about the way in which the data are entered into the primary databases

In order to better assess the quality of the data in the DWH LM&SP, it is necessary to understand how these data are gathered. First, a template was designed in order to standardise the documentation of this information. Next, this dossier was filled in after consulting with the relevant institutions for each source. This documentation is available to users via the CBSS website.

T1c: integrate existing and new documentation into the Content Management System, which is yet to be further developed

First, in consultation with Smals, the existing and new documentation were integrated into beConnected (an Alfresco application) in 2011. This online platform streamlines the procedure (= workflow) so that changes are applied to the documentation (content changes – translations – publication). Since then, this platform has been used to perform all updates and changes to the

documentation. Researchers have access to beConnected via an eID card reader.
Adjustments were also made to the existing documentation in terms of both form and content (file names, numbering of variables and codes, standardisation of tables, etc.) in preparation for the integration of the metadata into the Content Management System (CMS).
The functional development of the CMS began during the previous DW-AMSB project. However, the further development of the CMS was delayed during DOCDWH, though all documentation is currently uploaded to the system. The research team tested the CMS thoroughly for consistency, completeness, user-friendliness, etc. Once the latest bugs have been fixed by Smals, the CMS will be presented to the User Group (this is scheduled for the second quarter of 2014). This system will replace the Word documents that are available via the CBSS website. The CMS makes it possible for users to consult all documentation concerning the DWH LM&SP in a user-friendly way. Variables can be searched not only by institution, but also by source, topic, expiry date or alphabetically. It is also possible to consult the different levels of documentation (files, variables, codes, etc.) in a direct and interactive way.
All of this information that is currently available via the CBSS website has been integrated into the CMS and will be consultable via this application.

T2a: prepare documentation for the data from the FPS SS on disabled persons
The dossiers relating to the data of the FPS SS were created, validated by the FPS SS and published on the CBSS website. These dossiers have already undergone a number of revisions.

T2b: prepare documentation for the data of the NSSO on local places of business
The NSSO dossiers were supplemented with data from StatbaseD that have been available in the DWH LM&SP since 2007. StatbaseD has since contained information on the local place of business where the workers are employed. After being validated by the NSSO, the dossiers were published on the CBSS website.

T2c: prepare documentation for the data of the PPS SI on activation measures within the ESF framework
The PPS SI dossiers were supplemented with the description of the data on activation measures taken within the framework of ESF. After being validated by the PPS SI, the dossiers were published on the CBSS website.

T2d: prepare documentation for the data of other, potentially new primary and secondary databases
The career database - employees, which in the past had been supplied by CIMIRe, has been supplied by Sigedis since 2010. At that time, the career data for the period prior to 1990 were integrated into the DWH LM&SP. The documentation for these data was prepared, validated by Sigedis and published on the CBSS website. These dossiers have already undergone a number of revisions.
Documentation was also created for the new data in the DWH LM&SP. These are origin data, income variables and variables related to the LWI indicator.
Consultations took place with the National Annual Holiday Office (RJV) in connection with the integration of their data into the DWH LM&SP. An inventory was drawn up using the available and usable data. The CBSS has submitted a formal request to the RJV for the integration of these data.
Finally, the institution from which data on outbound cross-border labour are available was identified. Because an indication of outbound cross-border labour is available at the NIC, the CBSS has submitted a formal request to them for the integration of these data. These data are already available in the DWH LM&SP.

T3a: conduct an analysis to assess the comparability of the contents of the DWH LM&SP with those of the LFS (socio-economic position based on ILO criteria)
In a memorandum, the extent to which indicators for national and international reporting could also be calculated on the basis of the DWH LM&SP was determined. As a policy area, the EU 2020 strategy was chosen, specifically, the Work Intensity at Household Level. The EU-SILC

(and not the LFS) was used as a reference for this, which is customary for Work Intensity at the European level. The rules for the design of this indicator were presented to the User Group. This design was refined in consultation with the Directorate General Statistics and Economic Information (DGSEI) and the relevant data suppliers (NSSO, NSSOPLG and NISSE). Next, a data request was submitted in order to test the LWI indicator. Finally, the documentation was prepared so that the data would be available to users of the DWH LM&SP. This documentation includes the description of the variables and a methodological memorandum on the design of the variables. In addition, the first results are discussed in a contribution to the book entitled 'Het Datawarehouse, een duizendpoot! Perspectieven van het Datawarehouse Arbeidsmarkt en Sociale Bescherming - Le Datawarehouse, un mille-pattes! Perspectives du Datawarehouse Marché du Travail et Protection Sociale' (*The Datawarehouse, A Centipede! Perspectives of the Datawarehouse Labour Market and Social Protection*) (cf. below).

T3b: conduct an analysis to assess the comparability of the contents of the DWH LM&SP with those of the EU-SILC (income and poverty position based on EU-SILC criteria)
A comparison between the income concept of the DWH LM&SP and the income concept based on the EU-SILC is described in a contribution to the book entitled 'Het Datawarehouse, een duizendpoot! Perspectieven van het Datawarehouse Arbeidsmarkt en Sociale Bescherming - Le Datawarehouse, un mille-pattes! Perspectives du Datawarehouse Marché du Travail et Protection Sociale' (*The Data Warehouse, A Centipede! Perspectives of the Data Warehouse Labour Market and Social Protection*) (cf. below).

T4a: devise a framework for inventorying the most important policy measures impacting the domains, variables and codes of the DWH LM&SP
A framework was devised for inventorying the most important policy measures impacting the populations, sources, variables and codes of the DWH LM&SP. This framework was presented to the User Group and adjusted based on feedback from the users.

T4b: in a test phase, complete the framework for the NSSO, NEO and RVP for the years 2003-2009
The created framework was completed for the NSSO, NEO, FAO and RVP for the 2005-2010 period. During the test phase, the framework was adjusted based on the findings of the researchers and on feedback from the User Group and the relevant institutions. The legal dossiers from the FAO, NEO, NSSO and RVP were validated by the relevant institutions.
Due to the availability of sources and the usefulness to the user, it was decided to create this legal dossier for the 2005-2010 period instead of for the 2003-2009 period.
Following the evaluation of this test phase in the User Group, it was decided that it was neither worthwhile nor expedient to create these dossiers for the other sources. The legal dossiers of the NEO, RVP and NSSO are nonetheless available to the users via the CBSS website.

T5a: develop new derived variables for origin
Users were informed and consulted during a thematic User Group meeting related to the research into the socio-economic position of persons of foreign origin. Following consultation within the User Group, a procedure was launched at the initiative of the CBSS for the structural integration of genealogical data from the Belgian National Register into the DWH LM&SP. The CBSS submitted a request to the National Register to obtain these data. The first application turned out to be inadequate, so the CBSS submitted an authorisation request to the Sectoral Committee of the National Register. The authorisation from the Sectoral Committee of the National Register was granted, and the data on origin were integrated into the DWH LM&SP (2008-2012 period, which is followed by an evaluation). The new variables are described in dossiers that are available via the CBSS website.
Derived variables were also developed for migration history and nationality history. These variables were discussed in the User Group and have already been entered into the web-based application for global figures.

T5b: develop new derived variables for career periods

In consultation with Sigedis and the RVP, five derived variables for career periods were defined (days worked, days worked converted to FTE, equivalent days, equivalent days converted to FTE and wages). These variables were discussed in the User Group and were then put together.

An initial exploratory meeting was held with the NISSE regarding the career data of self-employed persons. The data available at the NISSE (including the eClipz database) were inventoried at this time.

T5c: develop new derived variables for income

An income concept (gross and gross taxable) was developed on the basis of the DWH LM&SP. This concept contains all income derived from both work and benefits that is available in the DWH LM&SP. The income was defined for each supplying institution (13) and on an annual basis per individual. After being validated by the institutions, this design was presented to the User Group. This income concept was also tested, and the results were submitted to the relevant institutions. The variables are described in new dossiers and how they are formed up is described in a methodological memorandum. This documentation is available via the CBSS website.

An income concept was also developed which describes an 'available income' based on administrative data. This concept contains a combination of income available in the DWH LM&SP (non-taxable income) and income available at the Federal Public Service Finances – IPCAL (taxable income). This memorandum was drawn up in consultation with the Federal Public Service Finances.

T6a: design and develop web-based applications for socio-economic mobility

In consultation with Smals and after consulting the User Group, the web-based application for socio-economic mobility was developed (2012). This web-based application makes it possible to track the socio-economic changes of the population entered into the DWH LM&SP (over the long term, specifically by year) either looking forward or backward in time. The following parameters were entered into the application: gender, age, region, nomenclature and designation of death and labour regime. This web-based application is available via the CBSS website.

T6b: design and develop a web-based application for family composition

The content of the web-based application for family composition was designed and discussed in the User Group, and a needs analysis was drawn up by Smals. However, this web-based application is not yet available due to budgetary reasons.

T7: track the nomenclature of the socio-economic position

The nomenclature of the socio-economic position was updated on a permanent basis. The following changes were also made:

- The designation of position n371 was changed (new designation: 'incapacity for work known to the health insurance funds').
- New derived variables related to the nomenclature were added.
- New values for the variables of the NEO were integrated into the nomenclature.
- The nomenclature was adjusted based on changes made to the Pension Register (RVP).
- The population of the FAO was integrated into the nomenclature (n374 'incapacitated for work due to workplace accident') and derived variables were formed for this population.
- The population of the FPS SS was integrated into the nomenclature (n38 'person with a disabled persons allowance'). At the same time, new derived variables were created for this population.
- The leave of absence taken prior to retirement was identified (as from 2011) in the nomenclature and represented by way of a new position (n352 'pre-retirement leave of absence') and its associated derived variables.

The integration of new data into the DWH LM&SP and into the nomenclature reduced the size of the group of persons that is not identified with a nomenclature position (n4).

T8: deploy the expertise on the effective linkage of various Belgian statistical sources to the DWH LM&SP through collaboration with other competent authorities
The research team submitted two data requests aimed at linking the DWH LM&SP to data from other authorities. The first request involves a link with the EU-SILC (DGSEI), and the second involves a link with the IPCAL data from the Federal Public Service Finances. Neither request has thus far been implemented. As a result, these requests could not be analysed.

T9a: provide feedback in the annual User Group
See section entitled 'User Group'.

T9b: provide feedback in the working group on origin data
See T5a.

T9c: provide feedback in the working group on the development of the income concept on the basis of the DWH LM&SP
After discussing the income concept in the User Group, a working group was formed. This working group is still in its infancy, but it has set the goal of the further development of an 'available income' based on administrative data. The IPCAL data from the Federal Public Service Finances, among others, will be used for this.

T10a: provide methodological support to the 'pan-sectoral impact of Social Security benefits' working group
The questions from the 'pan-sectoral impact of Social Security benefits' working group were answered. At the same time, a number of variables were designed within the context of the 'pan-sectoral impact of Social Security benefits'. This information was produced in respect of the FPS SS in two documents: one related to family statuses and the other related to socio-economic statuses.

T11: organise an international expert seminar
The international seminar for the DOCDWH project took place on 15 September 2010 at the Faculty Club in Leuven. Two non-Belgian experts provided clarification on databases constructed on the basis of administrative data (Prof. Bart Bakker: The Social Statistical Database, Statistics Netherlands; and Pernille Stender: Register-Based Labour Force Statistics, Statistics Denmark).

Intermediate meetings

The research teams had meetings with the CBSS and the FPS SS on a monthly basis. In these meetings, the status of the task were checked and the work was followed up. The reports of these meetings can be consulted by the coordinator of the project.

Support Committee

At the outset of the project, it was decided in consultation with the Belgian Federal Programmatic Service Science Policy, the CBSS and the FPS SS to allow the support committee of the DOCDWH project to converge with the pre-existing User Group of the DWH LM&SP. This User Group assembles on an ad hoc basis at least once a year. All members of the User Group make use of the DWH LM&SP; these members are primarily (academic) researchers, policymakers and data suppliers.

The User Group assembled seven times over the course of the project. In addition to a status report regarding the DWH LM&SP (sources, basic and web-based applications and documentation), the following topics were covered at meetings:

- 28 June 2010: The main focus was the integration of data on origin into the DWH LM&SP.
- 15 September 2010: The second meeting of the User Group coincided with the international seminar ("Databases compiled on the basis of administrative registers"). Two international experts were invited to this meeting (Prof. Bart Bakker and Pernille Stender).
- 20 December 2010: During this meeting of the User Group, the emphasis was on the presentation of the income concept on the basis of the DWH LM&SP.
- 23 November 2011: This User Group meeting covered the LWI indicator on the basis of the DWH LM&SP and an analysis of healthcare professions.
- 28 June 2012: The first results of the LWI indicator on the basis of the DWH LM&SP were presented during this meeting of the User Group. This also included a status report on the income concept as well as the life cycle and labour market indicators.
- 26 September 2012: Topics discussed during this meeting of the User Group included the initial results based on the origin data from the DWH LM&SP, a proposal concerning origin variables and a classification for the entry of the origin data into the applications of the DWH LM&SP.
- 26 June 2013: The User Group meeting provided a status report on the new web-based application for family composition, the dossiers with regard to the primary databases, the legal dossiers, the career codes of the career database (wage-earners) and the biographical indicators.

Book about the DWH LM&SP

Upon the conclusion of the DOCDWH project, the idea occurred to publish a book in connection with the DWH LM&SP. This book provides an overview of the possibilities of the DWH LM&SP. Firstly, the book contains a (brief) historical summary and a description of the current status of the DWH LM&SP. Secondly, it covers a number of studies conducted on the basis of the DWH LM&SP. The description of these studies paints a good picture of the possibilities of the DWH LM&SP. In order to cover all of the facets of the DWH LM&SP, the book includes descriptions of as many different types of studies as possible. Next, a number of applications that make use of data from the DWH LM&SP are explained. And finally, the book includes a discussion of indicators (biographical events, the LWI indicator and the income concept) that can be calculated on the basis of the DWH LM&SP. Instead of writing all of the chapters independently, the decision was made to collaborate with partners.

The book was entitled 'Het Datawarehouse, een duizendpoot! Perspectieven van het Datawarehouse Arbeidsmarkt en Sociale Bescherming - Le Datawarehouse, un mille-pattes! Perspectives du Datawarehouse Marché du Travail et Protection Sociale' (*The Datawarehouse, A Centipede! Perspectives of the Data Warehouse Labour Market and Social Protection*) (editors: Hans Knapen, Stijn Braes, Thomas Ermans and Wim Herremans). This book is published by Academia Press as part of its 'Science and Society' series. This series publishes final reports of social-scientific research projects that are organised and funded by the Belgian Federal Programmatic Service Science Policy and which form part of the Agora programme, among others. As such, this book also qualifies as the final report of the DOCDWH project. The book was published in February 2014.

*

* *